
Structure, expression and chromosomal localization of Zfp-1, a murine zinc finger protein gene

Kamal Chowdhury*, Susanne Dietrich, Rudi Balling, Jean-Louis Guenet¹ and Peter Gruss

Max-Planck-Institute for Biophysical Chemistry, 3400 Göttingen, FRG and ¹Unité de Genetique de Mummifères, Institute Pasteur, Paris 75724, Cedex 15, France

Received September 20, 1989; Revised and Accepted November 16, 1989

EMBL accession no. X16493

ABSTRACT

Zinc finger proteins (Zfp) are encoded by a large family of genes present in many organisms including yeast and human. Some of them are transcriptional activators and bind specifically to DNA by zinc mediated folded structures commonly known as zinc fingers. The *Drosophila* Krüppel (Kr) is a segmentation gene and encodes a zinc finger protein. Using a probe from the finger domain of Kr, we have isolated a structurally related gene Zfp-1 from the mouse. In this paper, we report the complete nucleotide sequence of two cDNA clones and the amino acid sequence deduced from them. The putative Zfp-1 protein contains in addition to 7 zinc fingers, two helix-turn-helix motifs. During murine embryogenesis, the Zfp-1 was found to express at a peak level in day 12 embryos. The ubiquitously expressed Zfp-1 gene is located in the 16q region on mouse chromosome 8, between the uvomorulin and the tyrosine amino transferase genes.

INTRODUCTION

Isolation and characterization of genes and gene products represent the initial step in the long chain of events leading to the study and understanding of gene regulation. The knowledge acquired from such experiments are crucial for the comprehension of the mechanism of cellular differentiation and pattern formation. In non vertebrates, like the fruitfly *Drosophila melanogaster*, the molecular biology of development has been greatly facilitated by the availability of a large number of well defined mutants and the p-element mediated gene transfer method to rescue these mutants. In contrast, the progress in the field of vertebrate development is much slower mainly due to lack of these two factors.

One approach to circumvent this difficulty is to use available *Drosophila* developmental genes as probes to isolate homologous mouse genes having potentially similar or related developmental function. This approach has been very fruitful in the isolation of a great number of *Drosophila* homeo box, zinc finger and paired box related genes from various organisms (1,2). By virtue of their temporally and spatially restricted expression pattern, some of these genes are potentially good candidates for playing developmental roles in the vertebrates.

One of these genes, the *Drosophila* Krüppel (Kr) is a segmentation gene of the gap class (3). Homozygous mutation in this gene is lethal, because the embryos fail to develop the 3 contiguous thoracic and the first 5 abdominal segments (4). However, apart from segmentation, Kr gene is also required for the development of

malpighian tubules. The structure of the Kr gene has been determined and found to encode 5 zinc finger repeats (5). The zinc fingers represent a major DNA binding motif and was first recognized as such in the TFIIIA gene (6). Such repeats have also been detected in many other eucaryotic proteins from sources as diverse as yeast and human (7,8). On the basis of their DNA binding properties and expression pattern, some of these vertebrate genes may have specific functions. For example, besides Kr, the mouse Zfp-2 (mkr2) is a neuron specific gene (9), the TDF has been related to sex determination (10), the krox-20 is a serum stimulated gene (11), the NGF1-A is induced by nerve growth factor (12) and the human SP1 is a well characterized ubiquitous transcription factor(13).

In the light of these findings, we decided to pursue the studies with one of our previous isolates, Zfp-1 (mkr1), which was cloned by its homology to the Drosophila Kr gene, and shown to contain 7 zinc fingers in the genomic DNA sequence (14). In order to gain further insight into the structure and function of Zfp-1, we have isolated full length cDNA clones, analysed its expression pattern in the developing mouse embryo and adult tissues and determined the chromosomal localization of the Zfp-1 gene, because if its location on the chromosome can be correlated to any known mouse mutation, it will immediately provide us with a handle to study its functional role in the mouse.

In this paper, we present the complete nucleotide sequence of two Zfp-1 cDNA clones capable of encoding 2 slightly different proteins. Zfp-1 is expressed at the peak level in day 12 mouse embryo and is present in varying amounts in different adult tissues. Furthermore, Zfp-1 is located on mouse chromosome 8 between the Uvomorulin and tyrosine amino transferase genes.

MATERIALS AND METHODS

Screening of the cDNA libraries

The mouse fibroblast cDNA library was constructed by H. Okayama in a SV40-plasmid based eucaryotic expression vector using the mRNA isolated from the cell line MB66 MCA ACL6 (15), and kindly provided to us. The screening and hybridization was performed using a 562 bp Eco RI fragment (nt 920-1482) from the genomic Zfp-1(mkr1) phage as described (9). This probe contains both unique and finger encoding sequences (14). Filters were washed twice for 15 minutes each in 2XSSC, 0.1% SDS and once in 0.1XSSC, 0.1% SDS at 65°C, dried and exposed.

The 8.5 day p.c. mouse embryo cDNA library (16) in λ gt10 was constructed by K.Fahrner and provided to us by B. Hogan. This library was screened with the 269bp Pst I fragment (146-415) of the cDNA c1 as described above.

DNA sequencing

Sequencing was performed by the dideoxy method using the Sequenase kit from Stratagene. Regions which were difficult to read were sequenced again by the Maxam-Gilbert method (17).

RNA isolation and northern blotting

Total RNA was isolated from developing mouse embryos or from adult mouse tissues by the guanidium thiocyanate method (18). PolyA⁺ RNA was prepared from the total RNA by retention on oligo dT cellulose column. Northern blotting was performed after electrophoresis of the poly A⁺ RNA in a formaldehyde-agarose gel as described earlier(9). Hybridization was performed using a unique cDNA

fragment (nt 228-760). The gene screen filters were washed twice with 2XSSC, 1% SDS, and once with 0.1XSSC, 1% SDS at 60°C for 15 minutes each as described earlier (9).

Computer analysis

All analyses were done using the University of Wisconsin Genetics computer group software package. For the structural analysis of the putative Zfp-1 protein, the peptide structure and the plot structure programs were used.

Chromosome mapping

Genomic DNAs of *Mus spretus* SPE/Pas mice, backcross progeny of (C57BL/6 x SPE/Pas) F1 females x C57BL/6 males, provided to us by J. Guenet, and of C57BL/6 mice were digested with Eco RI and blotted to Hxbond N membrane. The filters were hybridized o/n at 42°C with the oligo- labelled Alu I fragment (nt 228-760) from the Zfp-1 cDNA c1 in 50% formamide, 1% SDS, 1M NaCl and 100ug/ml denatured salmon sperm DNA. The blots were washed under stringent conditions and exposed for 48 hours. The contribution of the *Mus spretus* and C57BL/6 Zfp-1 alleles in the F2 backcross animals was compared to and matched with that of known genetic markers.

RESULTS

Structure of the Zfp-1 c-DNAs

Zfp-1 (mkr1) is expressed differentially in the mouse teratocarcinoma stem cell line F9 as a broad band of 2.0 kb transcripts (14). Upon differentiation of F9 cells with retinoic acid and cAMP into parietal endoderm, its concentration decreases by about 2 - 5 fold, but nevertheless, remains easily detectable (14). This allowed us to isolate a Zfp-1 cDNA clone from a differentiated mouse fibroblast cDNA library. Using a probe containing both the non finger and finger region of the genomic Zfp-1 (λ -mkr1)(9) clone, a single positive clone was isolated (c1). The entire DNA sequence of this clone was determined and found to contain 1804 bps, including the 30 adenine residues at the 3' end. It became evident that the size of c1 was smaller than the size of the Zfp-1 transcript. To isolate a full length clone, we therefore decided to rescreen another cDNA library with a different DNA probe obtained from the 5' end of c1.

An 8.5 day p.c. mouse embryo cDNA library (kindly provided to us by K. Fahrner and B. Hogan) was screened with the new probe and the longest clone obtained (c3) was analysed further by DNA sequencing. The Fig.1 is a composite of the 2 cDNA sequences consisting together of 1961 bps. Assuming that the Zfp-1 mRNAs contain on the average 100-200 adenine residues as poly A tail at the 3' end, we consider that our Zfp-1 cDNA sequence is complete. Nevertheless, we cannot exclude the possibility that a few nucleotides are missing from the extreme 5' end.

The sequence of c1 starts at nucleotide (nt) 117 and ends with the poly A present at the 3' end. The clone c3 starts at nt 1 and is colinear with c1. Its exact 3' end has not been determined by sequencing. But restriction enzyme mapping data show that its structure is similar to c1 except for the length of the poly A tail. Remarkably, c1 lacks a stretch of 42 nucleotides from the coding region of Zfp-1, shown as an insert between nt 282 and 283 in fig.1. This insertion of 42 nts (encoding 14 amino acids) in c3 is in frame with the protein coding region and is flanked by the direct repeat CCCAGG. The entire sequence has only one long open reading frame starting with 2 Methionines at nt 271 which is also in frame

1 AGAAGTACTG GATCCCCCGG GCTCCAGGAA TTCATGCTCT ACCTGCAAGG CTGGGTTGGG GTGAAGGCAC AGAAGCACAA AGGCACAGCT
91 GGAGGGGCTT AGCTTCACT CACAGCTCCG AGGACTCTGG GACCGCCCG TCGCTCCACA AGCTGCAAGT TTCCGCAAAA GCCTAGCCCG
181 TCGCCCTTCC CCCACCTGCA GTCCGCAAGT GGTGACAGA GCTCCCTTCT CCGCGAGCT CTGTAACCCCG AACCCGCGAGT GGAGGCTAGA

CAG GGA TCC GTT TCA TTC ACT GAT GTA ACT GTG GAC TTC ACC
Gln Gly Ser Val Ser Phe Thr Asp Val Thr Val Asp Phe Thr

271 CCCCTCCAGC TGTGCTTTTG TGCTTCTGTG CTTTCACCCC AACCCAGAAA ATG ATG GGG TCC CAG GAG GAG TGG GAG CAA CTG
1 Met Met Gly Ser Gln Glu Glu Trp Glu Gln Leu

354 GAC CCC TCG CAG AGG ATT CTG TAC ATG GAT GTA ATG CTG GAG AAC TAC AGC AAC TTA CTT TCA GTG GAA GTG TGG
12 Asp Pro Ser Gln Arg Ile Leu Tyr Met Asp Val Met Leu Glu Asn Tyr Ser Asn Leu Leu Ser Val Glu Val Trp

429 AAG GCT GAT GGC CAG GTG GAG AGG GAC CCC AGA GAC CTG CAG AGG CAG GTG GGG TCA CTT ACA ACC ATC AAG AAC
37 Lys Ala Asp Gly Gln Val Glu Arg Asp Pro Arg Asp Leu Gln Arg Gln Val Gly Ser Leu Thr Thr Ile Lys Asn

504 CAA CCA CCA ACT GAA GAA AGA GGC AGT CGT TTT GGA AAG ACA TTA ACT CTG AAC ACA GAC TTT GTT TCA TTA AGA
62 Gln Pro Pro Thr Glu Glu Arg Gly Ser Arg Phe Gly Lys Thr Leu Thr Leu Asn Thr Asp Phe Val Ser Leu Arg

579 CAA GTT CCC TAT AAA TAT GAT TTA TAT GAA AAA ACT TTG AAG TAC AAT TCA GAC TTA CTT AGC AGT AGA AAC TGT
87 Gln Val Pro Tyr Lys Tyr Asp Leu Tyr Glu Lys Thr Leu Lys Tyr Asn Ser Asp Leu Leu Ser Ser Arg Asn Cys

634 GTA AGA AAG AAA GGA GAC GGG TGT GGT GGA TTT GGG GAA CCA CTT CTG TAC CTG AAG CAA GAG AAA CCC CAC GCT
112 Val Arg Lys Lys Gly Asp Gly Cys Gly Gly Phe Gly Glu Pro Leu Leu Tyr Leu Lys Gln Glu Lys Pro His Ala

729 GGA CTG GAA TAC TCA GAA TAT AAT GGA AAC GGA AGG GCT CTC AGC CAT AAG GAT GCC ATC TTT AAA CAT CGG AAA
137 Gly Leu Glu Tyr Ser Glu Tyr Asn Gly Asn Gly Arg Ala Leu Ser His Lys Asp Ala Ile Phe Lys His Arg Lys

804 ATT AAG AGC TTG GTT CAG CCT TTT GTT TGT AAC TAC TGT GAC AAG ACC TTC TCC TTC AAG TCG CTG CTC GTT AGC
162 Ile Lys Ser Leu Val Gln Pro Phe Val Cys Asn Tyr Cys Asp Lys Thr Phe Ser Phe Lys Ser Leu Leu Val Ser

879 CAC AAG CGG ATC CAC ACG GGC GAG AAA CCC TAT GAG TGT GAT GTG TGC CAG AAA ACC TTC TCC CAT AAG GCC AAC
187 His Lys Arg Ile His Thr Gly Glu Lys Pro Tyr Glu Cys Asp Val Cys Gln Lys Thr Phe Ser His Lys Ala Asn

954 CTC ATC AAA CAT CAG AGA ATT CAC ACC GGA GAG AAG CCC TTT GAA TGT CCC GAG TGT GGG AAG GCC TTT ACC CAC
212 Leu Ile Lys His Gln Arg Ile His Thr Gly Glu Lys Pro Phe Glu Cys Pro Glu Cys Gly Lys Ala Phe Thr His

1029 CAG TCA AAC CTC ATC GTT CAC CAG AGA GCG CAC ATG GAG AAG AAG CCA TAT GGA TGC AGC GAG TGT GGG AAG ACA
237 Gln Ser Asn Leu Ile Val His Gln Arg Ala His Met Glu Lys Lys Pro Tyr Gly Cys Ser Glu Cys Gly Lys Thr

1104 TTC GCT CAG AAG TTT GAA CTC ACC ACA CAC CAG AGG ATT CAC ACA GGA GAG CGG CCC TAC GAG TGT AAC GAA TGC
262 Phe Ala Gln Lys Phe Glu Leu Thr Thr His Gln Arg Ile His Thr Gly Glu Arg Pro Tyr Glu Cys Asn Glu Cys

1179 GCA AAA ACC TTC TTT AAG AAA TCA AAC CTC ATC ATA CAT CAG AAG ATT CAC ACC GGC GAG AAA CGC TAC GAG TGC
287 Ala Lys Thr Phe Phe Lys Lys Ser Asn Leu Ile Ile His Gln Lys Ile His Thr Gly Glu Lys Arg Tyr Glu Cys

1254 AGC GAG TGC GGG AAG TCC TTC ATC CAG AAC TCG CAG CTC ATC ATC CAC AGG AGG ACT CAC ACT GGC GAG AAG CCC
312 Ser Glu Cys Gly Lys Ser Phe Ile Gln Asn Ser Gln Leu Ile Ile His Arg Arg Thr His Thr Gly Glu Lys Pro

1329 TAC GAG TGC ACC GAG TGT GGC AAG ACC TTC AGC CAG AGG TCA ACC CTG AGG CTG CAC CTG AGA ATC CAC ACC GGA
337 Tyr Glu Cys Thr Glu Cys Gly Lys Thr Phe Ser Gln Arg Ser Thr Leu Arg Leu His Leu Arg Ile His Thr Gly

1404 GAG AAA CCC TAC CAG TGT CCG CAG TGT GCG AAG CCT TCA GCA GGA AGT CCG CCT CAG CGT CCA CCA GAG GGT TCA
362 Glu Lys Pro Tyr Glu Cys Ala Glu Cys Gly Lys Pro Ser Ala Gly Ser Pro Pro Gln Arg Pro Pro Glu Gly Ser

1479 CAT GGC ATA GTC CTG AGG CTC AGC CAG GCT GCT GCA GAA AAC CTG TCA GGC CAG AAT TGT TCA GTG AAG GGG GAA
387 His Gly Ile Val Leu Arg Leu Arg Gln Ala Ala Ala Glu Asn Leu Ser Gly Gln Asn Ser Ser Val Lys Gly Glu

1554 GCC CTC ATG AAG ATG CTG AAG GAA CTT GGG AAA TTC TAT TGAGATGT AATCAATCTC AAAAAATTAA AAAAAAATA
412 Ala Leu Met Lys Met Leu Lys Glu Leu Gly Lys Phe Tyr

1631 AGCATACCTA ACAATATTAT ACTGAGGCTT CGAGAGATGG CTCGGCGGTT AAGACCACTT CCGGAGGTCC TGAGTTCAGT TCCGAGCAAC
1721 CACATGGTGG CTCACAACCA TGTGTAATGG CATCTATGCG CACCTTCTGG TGTCTGACGA CAGCAGCAGT GTACTCATAT AAAAAATA
1811 ATTTTAAAAA AAAAGATTAT ACTGAAAGTT ATATATATCCA GCTAAAGACT ACATATCCAA ATATATGTAT GTATCTCAGT ATGTGTAT
1901 TATCCAGATC AATTGTAAAT AGACATATCC AGCTCTGTTA AAAAAAATA AAAAAAATA AAAAAAATA

with the finger domain sequence. Upstream from these 2 ATGs, there are several stop codons in different reading frames. Both of these initiation codons correspond exactly to the eucaryotic consensus translational initiation site ANNATGA/G (19). There are several other Methionines located downstream (residue 20 and 23), but none of them correspond to the consensus initiation site. Two polyadenylation signals, AATAAA and AATAGA (20) are present at the 3' untranslated region located downstream from the termination codon at nt 1542. The complete Zfp-1 cDNA contains 270 and 388 nts at the 5' and 3' untranslated regions respectively.

Structure of the putative Zfp-1 protein

The conceptually translated larger Zfp-1 protein (c3) contains 438 amino acids and the smaller c1 protein consists of 424 amino acids (fig.2). Both open reading frames contain N- and C- terminal non finger sequences consisting of 168/182 and 59 amino acids respectively (fig.2). The finger domain located in the middle consists of 7 zinc finger repeats as shown by vertical lines in fig.1 and also separately in fig2. Each finger is 28 amino acids long and has the consensus sequence: YECXECGKXFXXXSLXXHXXXHTGEKP. The first 8 amino acids present in the C- terminal non finger domain could represent a partial finger sequence. When the cDNA derived protein sequence was compared to that encoded by the published partial genomic sequence (13), the first 4 amino acids from the genomic clone did not match with the cDNA derived protein sequence. Careful reexamination of the genomic sequence revealed a missing nucleotide "T" at position 12. All nucleic acid and peptide sequence downstream from this nucleotide are identical in both the cDNA and the genomic sequence. This alteration of the genomic sequence do not change the cDNA and the putative protein sequence derived from it.

The structure of the putative Zfp-1 protein was analysed, and to our surprise, in addition to the potential zinc and DNA binding finger motif, two putative helix-turn-helix motifs were also detected by both Chou- Fasman (21) and Garnier-Osguthorpe-Robson (22) methods (fig.3). The first motif encompassing 48 amino acid residues between 120-168 is located immediately upstream from the zinc finger motif. The second motif encompassing 34 amino acids is located at the extreme carboxy terminal end between residues 390 and 424. Both putative helix-turn-helix domains contain about 30% DNA binding amino acids (K,R,H,T,Q,N)(23). Calculation of the surface probability (24) of the entire Zfp-1 protein revealed that the region containing motif 1 has the highest value as compared to the other portions of the protein (data not shown). However, there are two potential glycosylation sites (NLS and NSS) located in the region of motif II. For this reason, this motif is not being shown in fig.3.

Figure 1. Nucleotide sequence of the Zfp-1 cDNA and the deduced amino acid sequence of the protein. The nucleotide sequence is a composite of 2 cDNAs. The longer clone c3 starts at nt. 1 and the shorter clone c1 starts at nucleotide 117. The 42 nt. insert located between nt. 282 and 283, is absent in c1. The finger domain is bracketed and individual fingers are separated by vertical lines. Potential glycosylation sites are boxed.

MMGS (QGSVSFTDVTVDFT) GEEWEQLDPSQRILYMDVML
 ENYSNLLSVEVWKADGQVERDPRDLQRQVGSLTITKNQPP
 TEERGSRFGKTLTLNTDFVSLRQVPYKYDLYEKTLYNSD
 LLSSRNCVRKKGDCGGFGEPLLYLKQEKPHAGLEYSWEY
 NGNGRALSHKDAIFKHKRIKSLVQP-
 FV C NY C DKT **F** SFKSL **L** VS **H** KRI **H** TGEKP
 YE C DV C QKT **F** SHKAN **L** IK **H** QRI **H** TGEKP
 FE C PE C GKA **F** THQSN **L** IV **H** QRA **H** MEKKP
 YG C SE C GKT **F** AQKFE **L** TT **H** QRI **H** TGEKR
 YE C NE C AKT **F** FKKS **L** II **H** QKI **H** TGEKR
 YE C SE C GKS **F** IQNSQ **L** II **H** RRT **H** TGEKP
 YE C TE C GKT **F** SQRST **L** RL **H** LRI **H** TGEKP
 YE C AE C GKPSAGSPQRPPEGSHGIVLRLRQAAAENL
 SGQNSSVKGEALMKMLKELGKFY*

Figure 2. Deduced amino acid sequence of the Zfp-1 protein, shown in one letter code. For the sake of clarity, the finger domain with individual fingers are shown separately. The conserved residues are in bold.

Comparison of the Zfp-1 finger domain sequence with that of other vertebrate multifinger proteins revealed that the Zfp-1 is structurally more closely related to other mouse Zfps (25) than to human TDF (10), SP1 (13), Zfp-7/GLI(27) or to Xenopus TFIIIA (28). This becomes more evident when one observes the H-C link sequence (29), conserved and present between individual fingers. We have compared 179 different individual fingers from published sequences. In no case we could detect greater than 60% homology. This indicates that although the Kr/TFIIIA like finger sequences are structurally similar, their actual sequence vary greatly. Interestingly, in spite of the species difference, the homology in the finger

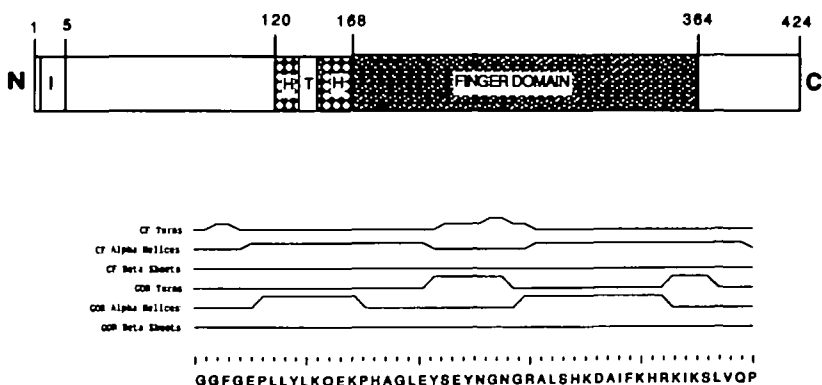


Figure 3. Schematic representation of the various domains of the Zfp-1 protein. I = 14 amino acid insert; HTH represents the helix turn helix motif and the non finger domains are shown as empty boxes. The lower panel shows the Chou - Fasman and the Garnier-Robson structure prediction of the HTH motif whose sequence is shown below.

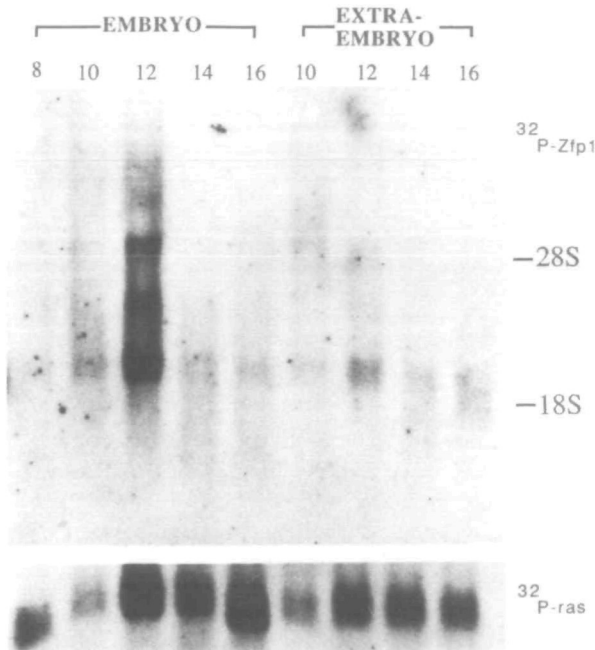


Figure 4. Expression of the Zfp-1 sequence in the developing mouse embryos as detected by northern blotting. Poly A⁺ RNAs were isolated separately from embryonic and extraembryonic tissues, electrophoresed, blotted onto Gene screen plus and hybridized as described in materials and methods. After removal of the probe, the same blot was hybridized to a c-Ha ras probe (lower panel).

regions of Zfp-1 and the *Drosophila* Kr was relatively high. This was not surprising to us, since the original genomic Zfp-1 clone was isolated by its homology to Kr.

The ubiquitous Zfp-1 transcripts are expressed early during embryogenesis

The *Drosophila* Kr is a segmentation gene of the gap class and expressed as a blastoderm-gastrulation stage-specific transcript in regions of the embryo affected by Kr mutants (4). Later, after completion of segmentation, it is also expressed in the developing nervous system. As a first step to find out the functional role of the structurally related Zfp-1 gene, it is necessary to analyze its expression pattern during embryogenesis and later after development, in various adult tissues. Earlier studies (13) have shown that it is expressed in F9 stem cells and can be detected as a broad 2.0 kb message in northern blots. The level of expression is decreased by 2-5 fold after differentiation.

The expression pattern of Zfp-1 during embryogenesis was studied by northern blotting of poly A⁺ RNA obtained from 8, 10, 12, and 14 day p.c. mouse embryos. RNAs were isolated separately from the embryonic and extraembryonic tissue and hybridized after electrophoresis and blotting to a probe from the 5' nonfinger region

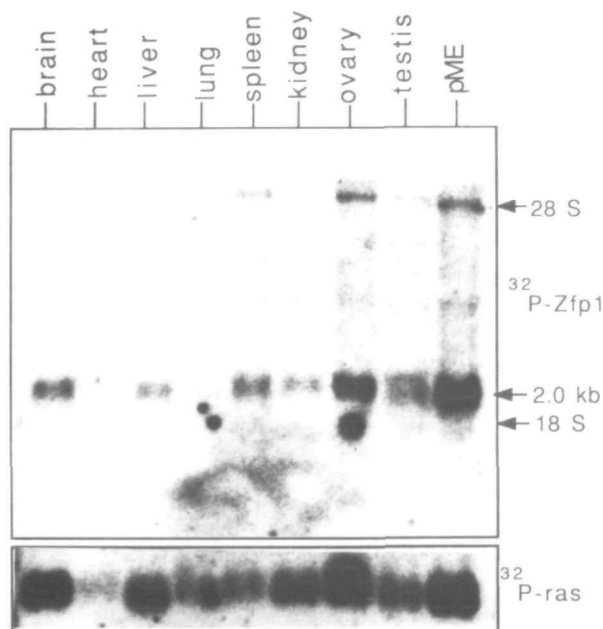


Figure 5. Expression of Zfp-1 sequence in various adult mouse tissues as detected by northern blotting. Poly A⁺ RNAs were isolated, blotted and hybridized to a Zfp-1 probe. The lower panel depicts the control hybridization of the same blot to a c - Ha ras probe. pME represents RNA isolated from primary mouse embryo fibroblast cells.

of Zfp-1 cDNA clone (nt 228-760). A broad band of about 2.0 kb transcript was already detectable in 8 day embryos and it continued to be expressed in all further developing embryonic and extraembryonic tissues (fig.4). Interestingly, day 12 embryos showed the highest level of expression. As an internal control of the amount of RNA loaded in each lane, the same blot was rehybridized to a Ha-ras probe, whose expression remains unaltered during embryogenesis (lower panel). A second minor band of about 4.8 kb was also detected in all lanes. Its expression paralleled that of the major transcript. Under the hybridization and washing (0.1XSSC, 60°C) conditions used, it is unlikely to be a different gene product. Alternatively, it could be a separate Zfp-1 transcript with a different initiation and/or termination sites or the non spliced transcript.

To analyze the fate and expression pattern of the Zfp-1 transcript(s), in the adult mouse, poly A⁺ RNAs were prepared from various tissues and analyzed again by northern blotting and hybridization. All tissues tested expressed the transcripts (fig.5), including thymus (data not shown). A tissue like an adult brain can be divided into several regions and contain different cell types. To find out if there is a spatially restricted distribution of expression among the various regions of the brain, a further northern blotting experiment was performed . In this experiment

RNA isolated from various regions of adult mouse brain were analyzed, but no obvious difference was noticed in the level of expression among these regions (data not shown). From these experiments we conclude that Zfp-1 is an ubiquitously expressed multifinger gene.

The chromosomal localization of the Zfp-1 Gene

To determine the localization of the Zfp-1 gene in the murine chromosome, we used the method of mouse interspecies backcross (reviewed by Guenet, 30). After Eco RI- digestion and hybridization with the 5' Alu I Fragment (nt 228-760) of the c1 cDNA, Restriction fragment length polymorphism between C57BL/6 genomic DNA and the inbred Mus spretus line SPE / Pas (maintained in the lab of J. Guenet) was detected. This RFLP was used to probe DNAs from a panel of backcross progeny of (C57BL/6 x SPE / Pas) F1 females x C57BL/6 males.

Matching the RFLPs to known genetic markers, we located Zfp-1 in the 16q region of the mouse chromosome 8. This segment is syntenic to human chromosome 16. Sixty of 61 recombinants revealed linkage between the Uvomorulin locus and Zfp1; 57 of 58 showed linkage between the Tyrosine amino transferase locus and Zfp-1. The linkage distance between Uvo and Zfp-1 is approximately 1,6+/-1,6 cM; between Tat and Zfp-1 1,7+/-1,7 cM. Since the Tyrosine amino transferase gene is located 5cM distal of the Uvomorulin locus, Zfp-1 maps between both genetic markers. There is no known mouse mutant at the Zfp-1 locus.

DISCUSSION

The DNA and metal binding zinc finger proteins (Zfps) can be grouped into two major classes. The first and smaller class contains the steroid receptor superfamily proteins (31, 32). Typical examples of this class are glucocorticoid, estrogen, and thyroid receptors. The single DNA and metal binding domain of these proteins has the following consensus amino acid sequence : $CX_2CX_{13}CX_2CX_{15-17}CX_5CX_9CX_2CX_4$ (X is variable amino acid). In presence of zinc, this domain can be folded to form two finger like structures separated from each other by 15-17 variable amino acids. The zinc mediated folding and DNA binding activity of these classes of proteins has been shown (33). Recently, by exchanging single amino acids in the first finger, it has been shown that the specificity for target DNA recognition lies in the two amino acids located between the 3rd and the 4th Cysteins (shown above in bold)(34, 35, 36). Simply by changing these two amino acids it is possible to alter the target specificity of the receptors. Several amino acids following the 4th Cysteine may also play a role in determining the specificity. The loop itself may only stabilize the binding. Furthermore, the second finger do not contribute to the DNA binding specificity and may be involved in protein-protein interaction. It should be mentioned however, that there are other amino acids involved in the DNA recognition that cannot be determined by these mutation studies. Interestingly, the target recognition region of the first finger has the potential of forming an α - helical structure similar to the ones present in the DNA binding helix-turn-helix motifs (37). A similar helical region consisting of the amino acid sequence $LX_2HX_{3-4}H$ has also been postulated to be present in the fingers of the second class, the Kr-TFIIIA family (38).

The second class is much larger and may include hundreds of genes in vertebrates (39) and usually contain multiple (up to 37) finger repeats. Each finger of this class has the following consensus amino acid sequence :

Y/FXCX₂CX₃FX₅LX₂HX₃HTGEKP. Due to the presence of the consensus Phe, Leu, and the linker sequence TGEKP, these fingers may possess a slightly different 3-dimensional structure than the steroid receptor fingers.

Zfp-1, isolated on the basis of its homology to the *Drosophila* Kr, belongs to the second family. By analogy to the receptor proteins, if the first finger of Zfp-1 should determine its target specificity, and if this specificity is determined by an identical region of the first finger (the putative α helical region LX₂HX₃₋₄H), then one would expect to find a great variability in the sequence of this region among different proteins of this family. We have compared the sequence of the first finger of Zfp-1 with that from all other published finger genes of this family. In no case we could detect more than 60% homology (data not shown). Furthermore, the sequence of the α -helical region was also variable in all of them. This would at least theoretically argue for them of having different target sites. Since the target sequence for TFIIIA is known, experiments similar to those done for the receptor proteins could be performed to answer these questions.

Secondary structure analysis of presumptive Zfp-1 protein revealed in addition, two putative helix-turn-helix DNA binding motifs by both Chou-Fasman and Garnier-Robson methods. Such motifs are present in some prokaryotic repressors and eucaryotic homeo domain proteins. It should be mentioned that homeo domains also contain an additional N-terminal helix, whose function in DNA binding is not known. Immediately upstream from this helix, there is a cluster of basic amino acids. This is also the case with the motif I in Zfp-1 (residue 113-115). In addition the second helix of this motif I contains 50% DNA binding amino acid residues. This is interesting in terms of the fact that, also the second helix of the homeo domain is the recognition helix and thought to interact with the major groove of target DNA. Comparison of the motif I with the antennapedia homeo domain revealed only 24% sequence identity and 40% sequence similarity. Considering the fact that the helix-turn-helix represents only a structural motif, a high degree of sequence conservation is not expected. Using similar approach, we could also detect similar motifs in the non finger regions of some finger proteins. For example, Kr (residue 50-100), Krox-20 (residue 20-80), and Egr-1 (residue 1-60, 265-310). Other finger proteins were not examined. We would however like to stress that, the mere detection of such motifs by computer analysis is by no means proof for their existence. Further characterization and possibly NMR studies of the protein will be necessary to clarify this point. Furthermore, even if such a motif exists, there is no guarantee that it will be involved in specific DNA interaction. For example, the helix-turn-helix motif present in the non specific DNA binding protein II from *B. stearothermophilus* is not involved in DNA binding (40). But considering the structural similarity, high number of DNA binding residues in the second helix and the high surface probability, we can not yet rule out the possibility that the helix-turn-helix motif of Zfp-1 may be involved in DNA binding.

The 2 Zfp-1 cDNAs isolated by us could encode 2 slightly different proteins varying by an insertion or deletion of 14 amino acids. Whether this difference is due to the different strains of mouse used as cDNA sources is not known. It would be necessary to characterize the corresponding regions from their genomic DNAs to answer this question. In this regard, the presence of a direct repeat (CCCAGG) flanking the insertion/deletion is noteworthy.

Zfp-1 is expressed like Kr, as a 2 closely migrating bands. Both transcripts are

expressed ubiquitously in the mouse throughout embryonic development and also in the adult tissues. Whether they are transcribed from two different promoters and/or have two different termination sites remains to be determined. By analogy to Kr and TFIIIA, Zfp-1 has the potential to be a zinc and DNA binding transcription factor. Another example of such an ubiquitous zinc finger containing transcription factor is the human SP1. The availability of full length cDNAs allows us to answer such questions.

ACKNOWLEDGEMENTS

We would like to thank H. Okayama and B.Hogan for providing us the cDNA libraries and Bernd Föhring for help with the computer work. We also thank G.Breier for providing the embryo northern blot and Thomas Schowalter for excellent technical assistance. This work was supported by DFG (Gr. 650/3-1) and the Max Planck Society.

*To whom correspondence should be addressed

REFERENCES

1. Scott, M., Tamkun, J., Hartzell, G. (1989). *Biochimica et Biophysica Acta*, 989, 25- 48
2. Dressler, G., and Gruss, P. (1988) *Trends in Genet.* 4(8), 214-219.
3. Nusslein-Vollhardt, C., and Wieschaus, E. (1980). *Nature*, 287, 785-801.
4. Knipple, D., Seifert, E., Rosenberg, U., Preiss, A., and Jäckle, H. (1985). *Nature*, 317, 40-44.
5. Rosenberg, U., Schröder, C., Preiss, A., Kienlin, A., Cote, S., Riede, E., and Jäckle, H. (1986). *Nature*, 319, 336-339.
6. Miller, J., MacLachlan, A. D., and Klug, A. (1985). *EMBO J.* , 4, 1609-1614.
7. Evans, R., and Hollenberg, S. (1988). *Cell*, 52, 1-3.
8. Rhodes, D., and Klug, A. (1988), in *Nucleic acid and molecular biology*. Ed. Eckstein, F., and Lilley, D., Springer Verlag, Berlin, Heidelberg, 1988.
9. Chowdhury, K., Dressler, G., Breier, G., Deutsch, U., and Gruss, P. (1988). *EMBO J.* , 7, 1345-1353.
10. Page, D., Mosher, R., Simpson, E., Fisher, E., Mardon G., Pollack, J., McGillivray, B., De la Chapelle, A., and Brown, L. (1987). *cell*, 51, 1091-1104.
11. Chavrier, P., Zerial, M., Lemaire, P., Almendral, J., Bravo, R., and Charnay, P., (1988) *EMBO J.* 7(1), 29-35.
12. Milbrandt, J. (1987). *Science*, 238-797-799.
13. Kadonaga, J., Carner, K., Masiaz, F., and Tjian, R. (1987). *Cell*, 51, 1079-1090.
14. Chowdhury, K., Deutsch, U., and Gruss, P. (1987). *Cell*, 48, 771-778.
15. Shih, C., Shilo, B., Goldfarb, M., Dannenberg, A., and Weinberg, B. (1979). *Proc. Natl. Acad. Sci. USA.* 76, 5714-5718.
16. Fahrner, K., Hogan, B., and Flavell, R. (1987) *EMBO J.* 6, 1265-1271.
17. Maxam, A., and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA.* 74, 560-564.
18. Chirgwin, J., Przbyla, A., McDonald, R., and Rutter, W. (1979). *Biochem.* 18, 5294 - 5299
19. Kozak, M. (1983). *Microbiol. Rev.* 47, 1-45.
20. Proudfoot, N., and Brownlee, G. (1976) *Nature*, 263, 211-214.

21. Chou, P. and Fassman, G. (1978) *Adv. in Enz.* 47, 45-48.
22. Garnier, J., Osguthorpe, D., and Robson, D. (1978). *J. Mol. Biol.* 120, 97-120.
23. Ohlendorf, D., and Mathews, B. (1983). *Ann. Rev. Biophys. Bioengng.*, 12, 259-284.
24. Emini, (1985) *J. Virol.* 55, 836-839.
25. Chavrier, P., Lemaire, P., Revelant, O., Bravo, R., and Chambay, P. (1988). *Mol. Cell. Biol.* 8(3), 1319-1326.
26. Chowdhury, K., Rohdewohld, H., and Gruss, P. (1988) *Nucl. acids. Res.* 16, 9995-10011
27. Kinzler, K., Ruppert, J., Bigner, S., and Vogelstein, B (1988). *Nature*, 332, 371-374
28. Brown, R. S., Sander, C., and Argos, P. (1985). *FEBS Lett.* 186, 271-274.
29. Schuh, R., Aicher, W., Gaul, U., Cote, S., Preiss, A., Maier, D., Seifert, E., Nauber, U., Schhröder, C., Kemler, R., and Jäckle, H. (1986) *Cell* 47, 1025-1032.
30. Guenet, J. (1986). in *Topics in microbiology and immunology*, 127, 109-130. Potter, M., Nadeu, J., and Cancro, M. Ed. Springer Verlag, Berlin.
31. Evans, R. (1988). *Science*, 240, 889-895.
32. Beato, M. (1989) *Cell*, 56, 335-344.
33. Freedman, L., Luisi, B., Korszun, Z., Basavappa, R., Sigler, P., and Yamamoto, K. (1988) *Nature*, 334, 543-546.
34. Mader, S., Kumar, V., de Verneuil, H., and Chambon, P. (1989). *Nature*, 338, 271-273.
35. Danielson, M., Hinck, L., and Ringold, G. (1989) *Cell*, 57, 1131-1138
36. Umesono, K., and Evans R. (1989) *Cell*, 57, 1139-1146.
37. Berg, J. (1989). *Cell*, 57, 1065-1068.
38. Brown, R., and Argos, P. (1986). *Nature*, 324, 215.
39. Bellefroid, E., Lecocq, P., Benhida, A., Poncelet, D., Belayew, A. and Martial J. (1989). *DNA*, 8(6), 377-387.
40. Tanaka, I., Appelt, K., Dijk, J., White, S., and Wilson, K. (1984). *Nature*, 310, 376-381